

0570  
0574

4/11



ENTERED

OIPE

RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/028,245

DATE: 05/13/2002

TIME: 09:51:18

Input Set : A:\GC700-SEQLIST.txt

Output Set: N:\CRF3\05102002\J028245.raw

4 <110> APPLICANT: Dunn-Coleman, Nigel  
5 Goedegebuur, Frits  
6 Ward, Michael  
7 Yao, Jian  
9 <120> TITLE OF INVENTION: EGVIII Endoglucanase and Nucleic Acids  
10 Encoding the Same  
12 <130> FILE REFERENCE: GC700  
14 <140> CURRENT APPLICATION NUMBER: US 10/028,245  
15 <141> CURRENT FILING DATE: 2001-12-18  
17 <160> NUMBER OF SEQ ID NOS: 4  
19 <170> SOFTWARE: FastSEQ for Windows Version 4.0  
21 <210> SEQ ID NO: 1  
22 <211> LENGTH: 1826  
23 <212> TYPE: DNA  
24 <213> ORGANISM: Trichoderma reesei  
26 <400> SEQUENCE: 1  
27 gtcgaccac gcgctcgttc attcttcttc cctctctctc cctctctctc ctcccccttct 60  
28 ccccatcac cgtcaccact ctctcattg ccgctctctc tgcgagccat gacgcagcat 120  
29 taacacacac tcgtttctgt tactctcgct gtcgtcggtt ctgctcggtg gcattctgct 180  
30 tagcactttg tttcgttctt cgttctcttt taatcgcgta tcttctgcaa tctgctgcca 240  
31 tttgttcgac taggtagtgg taatatacgg acagcttttt ttccctcgct caacacgctg 300  
32 acgtacaatt aatacaccat ctcgtaatac ggatataatc ctcggcctct tctgtgtgct 360  
33 tgtgcgacgc tctctgtttc tccctctcat tatgcgcgca acctcccttc tggccgccgc 420  
34 cttggccgtg gctggcgatg cctcgccgg caagatcaaa tatctgggcg tcgccattcc 480  
35 cggaatcgac tttggctgcg acatcgacgg cagctgtccg actgacacgt cgtctgtgcc 540  
36 cctgctgagc tacaaaggag gagatggcgc cggccagatg aagcatttcg ccgaagacga 600  
37 cggcctcaac gtctttcgca tatccgctac atggcagttt gtcctcaaca acacggtgga 660  
38 cggcaagctg gacgagctca actggggctc ctacaacaag gtcgtcaacg cctgtctcga 720  
39 gacgggcgcc tactgcatga ttgacatgca caactttgcc cgctacaacg gcggcatcat 780  
40 cggccaggga ggcgtgtcgg acgacatctt tgtcgacctc tgggtccaga tcgcaaagta 840  
41 ctacgaggac aacgacaaga tcatctttgg cctgatgaac gagccgcacg acctcgacat 900  
42 tgagatctgg gcgcagacgt gccaaaaggt cgtcactgcg atccgaaagg ccggcgccac 960  
43 ctgcgagatg atcctctctg ccggaaccaa ctttgccagc gtcgagacgt atgtgtccac 1020  
44 tggcagcgcg gaagccctcg gcaagattac gaaccggat ggaagcaccg atttgctgta 1080  
45 ctttgatgtc cacaagtatc tcgacatcaa caactccggg tcgcacgccg agtgaccac 1140  
46 agacaacgtc gacgccttca acgacttcgc ggactggctg aggcagaaca agcgccaggc 1200  
47 catcatctcc gaaacgggcg cgtccatgga accttcgtgc atgactgcct tctgcgccca 1260  
48 gaacaaggcc attagcgaaa acagcgacgt ctacattggc tttgtgggct ggggtgccgg 1320  
49 cagctttgac acgtcgtaca tcttgactct gactccctc ggcaagcccg gcaactacac 1380  
50 cgacaacaag ctcatgaacg agtgcattct ggaccagttt acctcgacg aaaagtaccg 1440  
51 tccaacaccc acctcaattt ccacagcggc ggaagagacg gccacggcga cagcaacctc 1500  
52 tgacggcgac gcgccatcca ctacgaagcc catctttagg gaagaaaccg cctctccccc 1560  
53 tccaatgct gttaccaagc cctcgcccga cagagcgac tcttccgacg acgacaagga 1620

## RAW SEQUENCE LISTING

DATE: 05/13/2002

PATENT APPLICATION: US/10/028,245

TIME: 09:51:18

Input Set : A:\GC700-SEQLIST.txt

Output Set: N:\CRF3\05102002\J028245.raw

```

54 ctcggcagca tctatgagtg cccagggctt gacaggcacg gtgctgttta ctgttgctgc 1680
55 ccttggctac atgctggtag cgttttgatg tttttttttt aatgagtttg tatacctaata 1740
56 gagcatgatt gagatgctac gtagtatata tgtcttttacg ggtacataag actagagcca 1800
57 tgttgtaatc aaaaaaaaaa aaaaaa 1826
59 <210> SEQ ID NO: 2
60 <211> LENGTH: 419
61 <212> TYPE: PRT
62 <213> ORGANISM: Trichoderma reesei
64 <400> SEQUENCE: 2
65 Gly Lys Ile Lys Tyr Leu Gly Val Ala Ile Pro Gly Ile Asp Phe Gly
66 1 5 10 15
67 Cys Asp Ile Asp Gly Ser Cys Pro Thr Asp Thr Ser Ser Val Pro Leu
68 20 25 30
69 Leu Ser Tyr Lys Gly Gly Asp Gly Ala Gly Gln Met Lys His Phe Ala
70 35 40 45
71 Glu Asp Asp Gly Leu Asn Val Phe Arg Ile Ser Ala Thr Trp Gln Phe
72 50 55 60
73 Val Leu Asn Asn Thr Val Asp Gly Lys Leu Asp Glu Leu Asn Trp Gly
74 65 70 75 80
75 Ser Tyr Asn Lys Val Val Asn Ala Cys Leu Glu Thr Gly Ala Tyr Cys
76 85 90 95
77 Met Ile Asp Met His Asn Phe Ala Arg Tyr Asn Gly Gly Ile Ile Gly
78 100 105 110
79 Gln Gly Gly Val Ser Asp Asp Ile Phe Val Asp Leu Trp Val Gln Ile
80 115 120 125
81 Ala Lys Tyr Tyr Glu Asp Asn Asp Lys Ile Ile Phe Gly Leu Met Asn
82 130 135 140
83 Glu Pro His Asp Leu Asp Ile Glu Ile Trp Ala Gln Thr Cys Gln Lys
84 145 150 155 160
85 Val Val Thr Ala Ile Arg Lys Ala Gly Ala Thr Ser Gln Met Ile Leu
86 165 170 175
87 Leu Pro Gly Thr Asn Phe Ala Ser Val Glu Thr Tyr Val Ser Thr Gly
88 180 185 190
89 Ser Ala Glu Ala Leu Gly Lys Ile Thr Asn Pro Asp Gly Ser Thr Asp
90 195 200 205
91 Leu Leu Tyr Phe Asp Val His Lys Tyr Leu Asp Ile Asn Asn Ser Gly
92 210 215 220
93 Ser His Ala Glu Cys Thr Thr Asp Asn Val Asp Ala Phe Asn Asp Phe
94 225 230 235 240
95 Ala Asp Trp Leu Arg Gln Asn Lys Arg Gln Ala Ile Ile Ser Glu Thr
96 245 250 255
97 Gly Ala Ser Met Glu Pro Ser Cys Met Thr Ala Phe Cys Ala Gln Asn
98 260 265 270
99 Lys Ala Ile Ser Glu Asn Ser Asp Val Tyr Ile Gly Phe Val Gly Trp
100 275 280 285
101 Gly Ala Gly Ser Phe Asp Thr Ser Tyr Ile Leu Thr Leu Thr Pro Leu
102 290 295 300
103 Gly Lys Pro Gly Asn Tyr Thr Asp Asn Lys Leu Met Asn Glu Cys Ile
104 305 310 315 320

```

## RAW SEQUENCE LISTING

DATE: 05/13/2002

PATENT APPLICATION: US/10/028,245

TIME: 09:51:18

Input Set : A:\GC700-SEQLIST.txt

Output Set: N:\CRF3\05102002\J028245.raw

105 Leu Asp Gln Phe Thr Leu Asp Glu Lys Tyr Arg Pro Thr Pro Thr Ser  
 106 325 330 335  
 107 Ile Ser Thr Ala Ala Glu Glu Thr Ala Thr Ala Thr Ser Asp  
 108 340 345 350  
 109 Gly Asp Ala Pro Ser Thr Thr Lys Pro Ile Phe Arg Glu Glu Thr Ala  
 110 355 360 365  
 111 Ser Pro Thr Pro Asn Ala Val Thr Lys Pro Ser Pro Asp Thr Ser Asp  
 112 370 375 380  
 113 Ser Ser Asp Asp Asp Lys Asp Ser Ala Ala Ser Met Ser Ala Gln Gly  
 114 385 390 395 400  
 115 Leu Thr Gly Thr Val Leu Phe Thr Val Ala Ala Leu Gly Tyr Met Leu  
 116 405 410 415  
 117 Val Ala Phe

120 &lt;210&gt; SEQ ID NO: 3

121 &lt;211&gt; LENGTH: 19

122 &lt;212&gt; TYPE: PRT

123 &lt;213&gt; ORGANISM: Trichoderma reesei

125 &lt;400&gt; SEQUENCE: 3

126 Met Arg Ala Thr Ser Leu Leu Ala Ala Ala Leu Ala Val Ala Gly Asp

127 1 5 10 15

128 Ala Leu Ala

131 &lt;210&gt; SEQ ID NO: 4

132 &lt;211&gt; LENGTH: 1317

133 &lt;212&gt; TYPE: DNA

134 &lt;213&gt; ORGANISM: Trichoderma reesei

136 &lt;400&gt; SEQUENCE: 4

137 atgcgcgcaa cctcccttct ggccgcccgc ttggccgtgg ctggcgatgc cctcgccggc 60  
 138 aagatcaa atctggggt cgccattccc ggaatcgact ttggctgcga catcgacggc 120  
 139 agctgtccga ctgacacgtc gtctgtgccc ctgctgagct acaaaggagg agatggcgcc 180  
 140 ggccagatga agcatttcgc cgaagacgac ggcctcaacg tctttcgcat atccgctaca 240  
 141 tggcagtttg tctcaacaa cacggtggac ggcaagctgg acgagctcaa ctggggctcc 300  
 142 tacaacaagg tcgtcaacgc ctgtctcgag acgggcccct actgcatgat tgacatgcac 360  
 143 aactttgccc gctacaacgg cggcatcatc ggccaggagg gcgtgtcgga cgacatcttt 420  
 144 gtcgacctct gggctccagat cgcaaagtac tacgaggaca acgacaagat catctttggc 480  
 145 ctgatgaacg agccgcacga cctcgacatt gagatctggg cgagacgtg ccaaaaggtc 540  
 146 gtcactgcga tccgaaaggc cggcgcccac tcgcagatga tctctctgcc cggaaccaac 600  
 147 tttgccagcg tcgagacgta tgtgtccact ggcagcgcgg aagccctcgg caagattacg 660  
 148 aaccgggatg gaagcaccga tttgtgttac tttgatgtcc acaagtatct cgacatcaac 720  
 149 aactccgggt cgcacgcga gtgcaccaca gacaacgtcg acgccttcaa cgacttcgcy 780  
 150 gactggctga ggcagaacaa gcgcaaggcc atcatctccg aaacgggcyg gtccatggaa 840  
 151 ccttcgtgca tgactgcctt ctgcgcccag aacaaggcca ttagcgaaaa cagcgacgtc 900  
 152 tacattggct ttgtgggctg ggggtgccggc agctttgaca cgtcgtacat cttgactctg 960  
 153 actcccctcg gcaagcccgg caactacacc gacaacaagc tcatgaacga gtgcattctg 1020  
 154 gaccagttta ccctcgacga aaagtaccgt ccaacaccca cctcaatttc cacagcgggc 1080  
 155 gaagagacgg ccacggcgac agcaacctct gacggcgacg cgccatccac tacgaagccc 1140  
 156 atctttaggg aagaaaccgc ctctccact cccaatgctg ttaccaagcc ctgcccgc 1200  
 157 acgagcgact cttccgacga cgacaaggac tcggcagcat ctatgagtgc ccagggcttg 1260  
 158 acaggcacgg tgctgtttac tgttgctgcc cttggctaca tgctggtagc gttttga 1317

VERIFICATION SUMMARY

PATENT APPLICATION: US/10/028,245

DATE: 05/13/2002

TIME: 09:51:19

Input Set : A:\GC700-SEQLIST.txt

Output Set: N:\CRF3\05102002\J028245.raw